IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 3, 2024

Fraud Detection in Online Product Review Systems via Heterogeneous Graph Transformer

[1]P MOUNIKA, [2]T.LOKESH

[1](Assistant Professor), MCA, **DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM ANDHRA PRADESH**

[2]MCA, scholar, **DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM ANDHRA PRADESH**

ABSTRACT

In online product review systems, users are allowed to submit reviews about their purchased items or services. However, fake reviews posted by fraudulent users often mislead consumers and bring losses to enterprises. Traditional fraud detection algorithm mainly utilizes rule-based methods, which is insufficient for the rich user interactions and graph-structured data. In recent years, graph-based methods have been proposed to handle this situation, but few prior works have noticed the camouflage fraudster's behavior and inconsistency heterogeneous nature. Existing methods have either not addressed these two problems or only partially, which results in poor performance. Alternatively, we propose a new model named Fraud Aware Heterogeneous Graph Transformer (FAHGT), to address camouflages and inconsistency problems in a unified manner. FAHGT adopts a type-aware feature mapping

ISSN 2347–3657

Volume 12, Issue 3, 2024

mechanism to handle heterogeneous graph data, then implementing various relation scoring methods to alleviate inconsistency and discover camouflage. Finally, the neighbors' features are aggregated together to build an informative representation. Experimental results on different types of real-world datasets demonstrate that FAHGT outperforms the state-of-the-art baselines.

1.INTRODUCTION

While the Internet has made many things easier for people to do, such as shop, connect with friends and family, and enjoy entertainment, it has also opened many doors for scammers. Scammers pose as regular users in order to release spam [1] or steal personal information, endangering the platforms' and users' interests. More than that, there are a plethora of connections linking various things on the Internet. This complex heterogeneous graph data is too much for traditional ML algorithms to manage. In order to find commonalities in the structure and traits of fraudsters, the present method involves modeling the data as a heterogeneous information network. Graph neural networks (GNNs) have previously been used in fraud detection domains such as product evaluation [2]–[5], mobile app distribution [6], cyber-crime identification [7], and financial services [8], [9] because of how well they learn graph representations. But most current GNN-based solutions only use homogeneous GNNs without considering the hidden node behaviors and the fact that the graph is heterogeneous. Numerous solutions have been suggested for this issue, which has garnered a lot of attention [4, 5, 10]. While CAREGNN [5] offered two more covert behaviors, Graph Consis [4] discovered three inconsistencies in fraud detection.

A brief summary of these issues would be: _

To hide their tracks, crowd workers have been shown to connect to trustworthy users or other benign entities, use special characters to mask fraudulent URLs, or use a generative language model to create fake reviews that don't belong to any particular domain [3, 6].

Inconsistency: Two people who have very different tastes in food or movies might end up linked by evaluating the same thing.

ISSN 2347–3657

Volume 12, Issue 3, 2024

Because of direct aggregation, GNNs are unable to differentiate between each user's own semantic pattern. Since dishonest people often write several fake reviews in a short amount of time, it stands to reason that if one person is suspicious, the other user should also be wary if they are related via a shared activity connection.

Many solutions have been suggested to deal with the two issues mentioned above. Graph Consis solves the inconsistency issue by calculating the similarity score between node embeddings, which does not differentiate between nodes of various kinds. CareGNN is an improvement over GNN-based fraud detectors that uses a relation-aware aggregator and a neighbor selection based on reinforcement learning to combat fraudsters who use camouflage. Due to its heterogeneous graph, its performance is still deficient. Here, we present the Fraud Aware Heterogeneous Graph Transformer (FAHGT), a system that uses label-aware neighbor selectors to combat camouflage and heterogeneous mutual attention to tackle inconsistencies. The "score head mechanism" unifies their implementation. Using a wide variety of real-world datasets, we prove that FAHGT is both efficient and effective. According to the findings of the experiments, FAHGT can outperform state-of-the-art GNNs in terms of KS and AUC.

2.LITERATURE SURVEY

Title:A REVIEW ON FAKE PRODUCT REVIEW DETECTION AND REMOVAL TECHNIQUES

Authors:RUTUJA B, GIRISH S

Description: The exponential growth of online shopping has necessitated effective review monitoring systems in e-commerce, where evaluating product reviews accurately is challenging due to the surge in online consumers and the prevalence of fake reviews. To address these challenges, the proposed method employs data mining techniques to analyze genuine product reviews, assisting both manufacturers and customers in distinguishing between positive and negative feedback. Users verify their identities using email accounts, receiving a summary of reviews based on chosen products and features. They can add reviews, which are automatically categorized as positive or negative, with fake accounts being identified and blacklisted. Regular updates ensure the reviews provide current information. The approach utilizes a combination of machine

ISSN 2347–3657

Volume 12, Issue 3, 2024

learning algorithms, such as BERT, Naive Bayes, SVMs, Random Forests, Neural Networks, K-means, Isolation Forests, One-Class SVMs, Matrix Factorization, Neural Collaborative Filtering, TF-IDF, and word embeddings (Word2Vec, GloVe), tailored to the project's requirements and implementation details. Experimental analysis and surveys demonstrate the approach's effectiveness and efficiency, highlighting its potential to enhance the online shopping experience.

Title:Detecting Fake Reviews in E-Commerce Platform

Authors: Arpitha, Ashwitha, Bhargavi, Deeksha, Sreedevi

Description: The exponential growth of online shopping has made online reviews a critical factor for customers, heavily influencing their purchasing decisions. However, the rise of fake reviews has become a major concern, as they can mislead potential buyers. To address this issue, a proposed solution involves an e-commerce platform that incorporates a review fraud detection system. This system uses Natural Language Processing (NLP) algorithms to analyze review features and sentiments, along with data science algorithms to detect fraudulent text. Within the platform, registered customers can view products, make purchases, and post reviews and ratings. The review detection model, utilizing the K-Nearest Neighbors (KNN) classification algorithm, classifies reviews as false or true based on customer inputs. This approach enhances the platform's credibility and helps ensure that customers can make informed purchasing decisions.

3. EXISTING SYSTEM

It is suggested that ChebNet [14] and GCN [15] may enhance efficiency via the use of approximation. In order to train GNNs on a spatial domain, GraphSAGE [16] takes a sample of a tree with each node serving as its root and calculates the hidden representation of the root by accumulating the representations of hidden nodes from the lowest level to the highest. Additional work on learning in the spatial realm is suggested by GAT [17] via the masked self-attention process, which involves calculating the varying relevance of neighbor nodes. For homogeneous graphs, none of these approaches work. A diverse graph with several kinds of items and interactions defies its straightforward application.

ISSN 2347–3657

Volume 12, Issue 3, 2024

There has been a proliferation of GNN-based heterogeneous approaches in the last few years. HAN [18], HAHE [19], and Deep- HGNN [20] use a handmade meta-path to turn a heterogeneous graph into several homogeneous graphs. Each of these graphs is then processed individually by GNN, and the result representations are aggregated by an attention mechanism. Nodes of the same object type are connected via the creation of meta-paths by Graph Inception [21]. As a first step, HetGNN [22] uses a random walk technique to sample a predetermined number of neighbors. Next, it uses a hierarchical aggregation approach to aggregate inside and across types. When applied to heterogeneous graphs, HGT [23] expands transformer architecture. Without taking domain knowledge into account, they aggregate based on attention ratings computed for all neighbors of a target node.

One approach to relation-aware graph fraud detectors is to construct several identical graphs using the original network's edge type information. Then, they may aggregate nodes based on type independently and join graphs depending on graph level configuration. Weighting parameters for various homogeneous subgraphs are learned using GEM [9]. While both Player2Vec [7] and SemiGNN [8] use attention mechanisms to aggregate features, SemiGNN goes a step further by using a structural loss to ensure that the node embeddings are homophilic. In order to aggregate diverse data in the graph, some works do it directly. As an example, in a user-review-item heterogeneous network, GAS [3] repeatedly updates the embeddings of each node type after learning a unique set of aggregators for distinct kinds of nodes.

Disadvantages

- In the existing work, the system did not implement Fraud Aware Heterogeneous Graph Transformer(FAHGT) to measure frauds exactly.
- This system is less performance due to lack of META RELATION SCORING.

3.1 PROPOED SYSTEM:

While calculating the similarity score across node embeddings can help GraphConsis deal with inconsistencies, it is unable to differentiate between nodes of various kinds. CareGNN is an improvement over GNN-based fraud detectors that uses a relation-

ISSN 2347–3657

Volume 12, Issue 3, 2024

aware aggregator and a neighbor selection based on reinforcement learning to combat fraudsters who use camouflage. Due to its heterogeneous graph, its performance is still deficient.

This study presents the Fraud Aware Heterogeneous Graph Transformer (FAHGT), a system that addresses inconsistencies and camouflage via the use of heterogeneous mutual attention and a label-aware neighbor selection. The "score head mechanism" unifies their implementation. Using a wide variety of real-world datasets, we prove that FAHGT is both efficient and effective. Based on the experimental data, FAHGT outperforms both state-of-the-art GNNs and fraud detectors that use GNNs in terms of KS and AUC

TheBenefits

The following is a synopsis of FAHGT's benefits.

• Heterogeneity: FAHGT can manage varied networks with different types of relationships and nodes automatically, without the need to build meta-paths by hand.

• Flexibility: FAHGT carefully chooses neighbors using real-world data and a noise graph. For feature aggregation, the chosen neighbors may be instructive, but for fraud detection, they may be dangerous.

• Efficient: FAHGT permits minimal computing complexity in connection scoring and feature aggregation using a parallelizable multi-head technique. The adaptable relation scoring technique introduced by FAHGT allows for the injection of domain knowledge. A relation's score between two nodes is limited by domain knowledge and derived from direct feature interaction.

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 3, 2024

4. OUTPUT SCREENS

Registration page:

Login Page:

Detection of Product Review Type:No Fraud Review

Detection of Product Review Type: Fraud Review

Admin Page:

Browse And Train Dataset:

View Detection Product Review Type:

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 3, 2024

Find Detection Product Review Type Ratio:

View Detection Product Review Type Ratio:

5. CONCLUSION

For the purpose of detecting fraudulent users in online review platforms, we present FAHGT, a new heterogeneous graph neural network. In order to deal with features that aren't consistent, we use heterogeneous mutual attention to build automated meta paths. As a means of detecting covert actions, we developed label aware scoring with the intention of excluding disruptive neighbors. The "score head mechanism" is a unifying method of combining two neural modules; each module contributes to the calculation of edge weights in the final feature aggregation. Results from experiments conducted on actual business datasets confirm that FAHGT has a very effective influence on fraud detection. Our model's stability and efficiency are further shown by the hyper-parameter sensitivity and visual examination. Overall, FAHGT provides state-of-the-art performance in most cases by relieving inconsistency and discovering concealment. Future plans include expanding our model's ability to handle data from dynamic graphs and integrating fraud detection into other domains, such financial services for loan default prediction or E-commerce for strong item recommendat

6.REFERENC

[1] A. Li, Z. Qin, R. Liu, Y. Yang, and D. Li, “Spam review detection withgraph convolutional networks,” in CIKM, 2019.

[2] Z. Liu, Y. Dou, P. S. Yu, Y. Deng, and H. Peng, “Alleviating the inconsistencyproblem of applying graph neural network to fraud detection,” in SIGIR, 2020.

ISSN 2347–3657

Volume 12, Issue 3, 2024

[3] Y. Dou, Z. Liu, L. Sun, Y. Deng, H. Peng, and P. S. Yu, "Enhancing graph

neural network-based fraud detectors against camouflaged fraudsters," inCIKM, 2020. [6] R. Wen, J. Wang, C. Wu, and J. Xiong, "Asa: Adversary situation awareness via heterogeneous graph convolutional networks," in WWW

Workshops, 2020.

[4] Y. Zhang, Y. Fan, Y. Ye, L. Zhao, and C. Shi, "Key player identification in underground forums over attributed heterogeneous information network

embedding framework," in CIKM, 2019.

[5] D. Wang, J. Lin, P. Cui, Q. Jia, Z. Wang, Y. Fang, Q. Yu, and J. Zhou, "A semi-supervised graph attentive network for fraud detection," in ICDM, 2019.

[6] Z. Liu, C. Chen, X. Yang, J. Zhou, X. Li, and L. Song, "Heterogeneous graph neural networks for malicious account detection," in CIKM, 2018.

[7] Y. Dou, G. Ma, P. S. Yu, and S. Xie, "Robust spammer detection by nashreinforcement learning," in KDD, 2020.

[8] P. Kaghazgaran, M. Alfifi, and J. Caverlee, "Wide-ranging review manipulation attacks: Model, empirical study, and countermeasures," in CIKM,

2019.

[9] Z. Zhang, P. Cui, andW. Zhu, "Deep learning on graphs: A survey," TKDE, 2020.

[10] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," arXiv preprint arXiv:1312.6203, 2013.